UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 03-40
Economics Series 14
October 2003

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624 98 75

AN EVALUATION OF AN ENTROPY BASED INDEX OF SEGREGATION *

Ricardo Mora[1] and Javier Ruiz-Castillo[2]

Abstract

This paper reviews the properties suggested in the methodological literature on the measurement of gender segregation by occupation. It is found that an index of segregation based on the entropy concept satisfies twelve basic axioms previously proposed in the single-dimensional case. This index can be expressed as the sum of a between-group and a within-group term in the two-dimensional case. In pair-wise comparisons, it can be meaningfully decomposed into three terms, one of which is independent of both the gender composition of the population and the population's distribution across occupations. Finally, it can be motivated as two different log-likelihood tests in nonparametric econometric models. Other existing measures of segregation either fail to satisfy one or more of the basic axioms, do not admit a between/within decomposition, have not been motivated from a statistical approach, or are based on more restricted econometric models.

Keywords: gender segregation measurement; axiomatic properties; econometric models.

JEL Classification: J16, J24.

[1] R. Mora, Departamento de Economía, Universidad Carlos III de Madrid; E.mail: ricmora@eco.uc3m.es. Phone: (34) 91 624 9576
[2] J. Ruiz-Castillo, Departamento de Economía, Universidad Carlos III de Madrid. E.mail: jrc@eco.uc3m.es. Phone: (34) 91 624 9588

* The authors acknowledge financial support from the Spanish DGI, Grants BEC2003-03943 and BEC2000-0173I.

I. INTRODUCTION

Social scientists have long been interested in the problem of segregation in the labor market by gender, that is, the tendency of men and women in the employment population to be differently distributed across occupations.[1] The information contained in the joint distribution of gender and occupation is usually summarized by means of numerical indices of segregation. In spite of the large volume of contributions, most of the proposed indices fall into the following three categories.

The first family of indices refers to those inspired by the Index of Dissimilarity, ID, first proposed in Duncan and Duncan (1955). The popularity of this index is based on its appealing interpretation as the proportion of male or female workers that would have to be removed without replacement in order to make every occupation contain the same gender mix exhibited by the labour force as a whole. This interpretation is at the core of the development of several variants of the index.[2] A second approach exploits the connection between the measurement of income inequality and the measurement of gender segregation viewed as the inequality in the distribution of the employed population across occupations. This is the case of indices inspired in the Gini index of income inequality, as well as the family of Atkinson's indices, the coefficient of variation or one of Theil's measures.[3] Finally, a structural approach to gender segregation measurement has been recently advocated under the argument that the

---

[1] The seminal article on (residential) segregation is Duncan and Duncan (1955). For recent contributions to gender segregation, see the special issues of the *Journal of Econometrics*, 1994, **61**(1), and *Demography*, 1998, **35**(4), as well as the treatise by Flückiger and Silber (1999).

[2] See Cortese *et al.* (1976), Moir and Selby Smith (1979), Lewis (1982), Karmel and MacLachlan (1988), Silber (1992), and Watts (1992). The index and its variants have become so dominant after the "index wars" (Peach, 1975), that concern has recently been voiced about a situation in which it is generally "assumed that sex segregation is simply whatever ID measures" (Grusky and Charles, 1998).

conventional practice of using a scalar index to describe gender segregation differences over time and/or across countries must be embedded in a testable model. This is the case of Charles (1992, 1998), Charles and Grusky (1995) and Grusky and Charles (1998), who propose a log-multiplicative model, or Kakwani (1994) who develops a procedure based on the F-distribution to test whether gender segregation has increased or decreased significantly within any two periods or across any two countries.

This paper defends the use of an index, I_E, based on the entropy concept used in information theory. It was first introduced in the segregation literature by Theil and Finizza (1971) and Fuchs (1975), and has recently been extended to the multidimensional case by Herranz *et al.* (2003) and Mora and Ruiz-Castillo (2003a).

Naturally, two segregation indices may show different trends in a given country, and may produce different country rankings in international comparisons.[4] Thus, the design of measures with desirable properties is a central methodological issue, and the merits of competing indices are regularly debated.[5] For our purposes, the properties of segregation indexes discussed in the literature can be classified into four groups. First, there is a number of basic desirable characteristics for the case in which gender segregation takes place along a single dimension, say occupation. Second, there is an important group of invariance axioms that serve two purposes: (i) to characterize relative (or size invariant) *versus* absolute

[3] See, *inter alia*, Duncan and Duncan (1955), Schwartz and Winship (1979), Butler (1987), Silber (1989a, 1989b), Hutchens (1991), and Flückiger and Silber (1999).

[4] For some evidence in this respect, see *inter alia* Jonung (1984), James and Tauber (1985), Karmel and Maclachlan (1988), Blackburn *et al.* (1993), and Flückiger and Silber (1999).

[5] See *inter alia*, the methodological contributions by James and Tauber (1985), Siltanen (1990), Hutchens (1991, 2001), Watts (1992, 1997a, 1998a, 1998b), Blackburn *et al.* (1993, 1995), Kakwani (1994), Charles (1992), Charles and Grusky (1995), Grusky and Charles (1998), and Flückiger and Silber (1999).

segregation measures; and (ii) to make precise what is meant by a margin-free index, that is, a segregation index that is independent from changes in the overall share of employment by gender (composition invariance), and from changes in the occupational structure (occupational invariance). Third, when segregation takes place along two dimensions, say educational level and occupation, it is useful that overall segregation can be expressed as the sum of two terms. The first term captures the *between-group* segregation induced by one of the classification variables, while the second term records the segregation induced by the second variable *within* the groups defined by the first one. Finally, since segregation measures are usually computed using sample observations, an additional desirable property for a measure of segregation is that it is embedded in a statistical framework that permits the testing of hypothesis on gender segregation in occupations.

This paper contributes to the existing literature in two ways. First, it evaluates the entropy based index I_E and shows that it satisfies twelve basic properties in the single-dimensional case, and that it is decomposable into a *between-group* and a *within-group* term in the two-dimensional case. In addition, it is shown that although I_E is neither composition nor occupational invariant, pairwise comparisons of the index can be decomposed so that one of the terms in the decomposition captures changes in the index which are unrelated to changes in the overall female share or the employment distribution across occupations.

In the second contribution to the existing literature, it is shown that the I_E index has two straightforward interpretations based on log-likelihood tests. The econometric models underpinning the tests encompass two alternative statistical notions of segregation. More specifically, it is shown that I_E is a monotonic transformation of the log-likelihood ratio test for the equality of both the male and female distributions and the male and female shares across occupations.

To our knowledge, other existing measures of segregation either fail to satisfy one or more of the basic axioms in the single-dimensional case, cannot be expressed as the sum of a between-group and a within-group term in the two-dimensional case, have not been motivated from a statistical approach, or are based on more restricted econometric models.

The rest of the paper contains four Sections. Section II reviews the main axioms discussed in the literature. The I_E index of segregation is presented in Section III, where its properties are also studied. Section IV is devoted to the statistical properties of I_E, while Section V offers some concluding comments.

II. BASIC AXIOMS

II. 1. The Single-dimensional Case. Notation

Assume an economy with J occupations, indexed by $j = 1,\ldots, J$. The usual data available in empirical situations can be organized into the following (3 x (J + 1)) array

$$\begin{matrix} F_1, F_2, \ldots, F_J & F & & \mathbf{f} & F \\ M_1, M_2, \ldots, M_J & M & = & \mathbf{m} & M \\ T_1, T_2, \ldots, T_J & T & & \mathbf{t} & T \end{matrix}$$

where $\mathbf{f} = (F_1, F_2,\ldots, F_J)$, $\mathbf{m} = (M_1, M_2,\ldots, M_J)$ and $\mathbf{t} = (T_1, T_2,\ldots, T_J) = (F_1 + M_1, F_2 + M_2,\ldots, F_J +$

M_J) are the (1 x J) vectors of females, males, and people, respectively, employed in each occupation, whereas $F = \Sigma_j F_j$, $M = \Sigma_j M_j$ and $T = \Sigma_j T_j$ are, respectively, the total number of females, males, and people in the economy.

For later reference, define three types of (1 x J) vectors. First, the vectors $\mathbf{s}^f = (s^f_1,\ldots, s^f_J) = (F_1/F,\ldots, F_J/F)$, $\mathbf{s}^m = (s^m_1,\ldots, s^m_J) = (M_1/M,\ldots, M_J/M)$ and $\mathbf{s}^t = (s^t_1,\ldots, s^t_J) = (T_1/T,\ldots, T_J/T)$, capturing the frequency distributions over occupations of females, males and people, respectively. Second, the vectors $\mathbf{w} = (w_1,\ldots, w_J) = (F_1/T_1,\ldots, F_J/T_J)$ and $(\mathbf{1} - \mathbf{w}) = (1 - w_1,\ldots, 1 - w_J) = (M_1/T_1,\ldots, M_J/T_J)$ of female and male shares in all occupations. Third, the vector of gender ratios $\mathbf{r} = (r_1,\ldots, r_J) = (F_1/M_1,\ldots, F_J/M_J)$. Finally, denote the overall female and male shares by $W = F/T$ and $(1 - W) = M/T$, respectively, and the overall gender ratio by $R = F/M$.

In many contexts, numerical indexes serve to summarize the degree of gender segregation prevailing in the entire economy, and provide a concise means of presenting the dominant trends that may be hidden in a detailed occupation by occupation study. For the sake of generality, a distribution of people across gender and occupations will be identified in the sequel by a 6-tuple ($\mathbf{f}$, F, $\mathbf{m}$, M, $\mathbf{t}$, T). Any scalar index of segregation, θ, can then be seen as a unique real non-negative valued function of ($\mathbf{f}$, F, $\mathbf{m}$, M, $\mathbf{t}$, T), $\theta = \theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T)$.[6]

A number of desirable properties for an index of segregation have been proposed, among others, by James and Taeuber (1985), Siltanen (1990), Kakwani (1994), and Hutchens

[6] Of course, this formal framework is equally well suited for the measurement of other segregation phenomena, such as the segregation exhibited by the distribution of black and white students over schools in a given school district.

(1991). These properties will be presented below as axioms. However, these axioms need not be considered all desirable at the same time. As in Kakwani (1994), the purpose here is not so much to justify them as to provide a framework for comparing various segregation indices.[7]

II. 2. A Notion of Occupational Gender Segregation and Basic Axioms

All notions of occupational gender segregation stem from an idea of association between gender and occupational category. In the majority of instances, segregation is said to exist when women and men are differently distributed across occupations that they are in employment overall, regardless of the nature of job allocation (Jonung, 1984). It is usually understood that an index of gender segregation $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T)$ based on this notion measures the extent to which the female and the male distributions differ across occupations. This is why some of the basic axioms presented in the sequel (in particular, A.1, and A.6 to A.9), as well as definition 1 will be couched in terms of the vectors $\boldsymbol{s}^{\mathrm{f}}$ and $\boldsymbol{s}^{\mathrm{m}}$.

Explicit in the calculation of any index is the specification of two counterfactual distributions that capture the ideas of complete integration and complete segregation. Within the above notion of occupational gender segregation, there is broad agreement on the meaning of what these two distributions should be.

Axiom 1: (*Complete Integration*, Kakwany 1994) Let $(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T)$ be such that $\mathbf{s}^{\mathrm{f}} = \mathbf{s}^{\mathrm{m}}$. Then $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = 0$. **o**

Notice that this relative notion of complete integration is not the only one within this approach. Chakravarty and Silber (1992) suggest an absolute (and stronger) notion of complete

[7] This approach can be contrasted to Hutchens (2001) and Chakravarty and Silber (1992), the only two studies in

integration, according to which there is no gender segregation if and only if $F_j = M_j$ for all j.[8]

Axiom 2: (*Complete Segregation*, Kakwany 1994) Let ($\mathbf{f}$, F, $\mathbf{m}$, M, $\mathbf{t}$, T) be so that F_j (M_j) > 0 implies M_j (F_j) = 0 for all j. Then $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = 1$. **o**

This axiom implies that the index should have a maximum value of unity when females and males are in separate occupations.

The next two axioms capture two different symmetry notions.

Axiom 3: (*Symmetry in Groups*, Kakwany 1994 and Hutchens 1991) Let $\mathbf{f'}$ and $\mathbf{m'}$ be two permutations of $\mathbf{f}$ and $\mathbf{m}$, respectively. Then $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = \theta(\mathbf{f'}, F, \mathbf{m'}, M, \mathbf{t'}, T)$. **o**

Axiom 4: (*Symmetry in Types*, Kakwany 1994 and Hutchens 2001) $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = \theta(\mathbf{m}, M, \mathbf{f}, F, \mathbf{t}, T)$. **o**

For the next axioms, it is useful to introduce the following:

Definition 1: An occupation j is *female dominated* if and only if $s^{f}_{j} > s^{m}_{j}$. **o**

Axiom 5: (*Weak Principle of Transfers*, James and Tauber, 1985, Kakwani 1994) If there is a small shift of the female labor force from a female- (male-) dominated occupation to a male- (female-) dominated occupation, the segregation index must decrease. **o**

Siltanen (1990) and Watts (1992) propose a somewhat stronger condition than A.5, which is also closely related to the following:

Axiom 6: (*Movement between Groups*, Hutchens 1991) Let $M'_h = M_h = M'_j = M_j$ for any h,

---

the segregation literature that attempt an axiomatic characterization of specific numerical measures.

[8] In the *marginal matching* approach advocated by Blackburn *et al.* (1993, 1995), occupational gender segregation is "the relationship between gendering of occupations and the sex of the workers, measuring the tendency for men and women to work in different occupations". In this context, zero segregation is defined differently from A. 1. For a critical assessment of this approach, which lies beyond this paper's scope, see Watts (1994, 1997b).

j. Assume that there are two occupations i and k such that: (a) $(s^f_i/s^m_i) < (s^f_k/s^m_k)$, (b) $F'_i = F_i$ - d and $F'_k = F_k$ + d, for $0 < d \leq F_i$, and (c) $F'_j = F_j$ for any $j \neq i, k$. Then $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) < \theta(\mathbf{f'}, F, \mathbf{m'}, M, \mathbf{t}, T)$. **o**

This disequalizing movement is similar to a regressive transfer in the income inequality literature. It reduces the presence of women in a given occupation, and increases it in an occupation that originally has a higher ratio of women to men. Therefore, A.6 is closely related to the Pigou-Dalton principle in the income inequality literature.

In the context of residential segregation, Zoloth (1976) introduced the notion of *diminishing payoffs to desegregation* as a useful property from a policy point of view, arguing that the cost of additional desegregation rises with the level of desegregation already achieved. This notion is analogous to the property of *decreasing returns of inequality in proximity* in Kolm (1999), or the *transfer sensitivity* property in Shorrocks and Foster (1987) in the income inequality literature. This idea can be formulated as a stronger condition than A.6:

Axiom 7: (*Increasing Returns to a Movement Between Groups*) Let $M''_h = M'_h = M_h = M''_j = M'_j = M_j$ for any h, j. Assume that there are two occupations i and k such that: (a) $(s^f_i/s^m_i) < (s^f_k/s^m_k)$, (b) $F''_i = F'_i$ – d, $F''_k = F'_k$ + d, $F'_i = F_i$ - d and $F'_k = F_k$ + d, for $0 < 2d \leq F_i$, and (c) $F''_j = F'_j = F_j$ for any $j \neq i, k$. Then $[\theta(\mathbf{f''}, F, \mathbf{m''}, M, \mathbf{t}, T) - \theta(\mathbf{f'}, F, \mathbf{m'}, M, \mathbf{t}, T)] > [\theta(\mathbf{f'}, F, \mathbf{m'}, M, \mathbf{t}, T) - \theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T)] > 0$. **o**

The following two axioms impose value judgements on how the size of the occupations should influence the index of segregation.

Axiom 8: (Kakwani 1994) If i and k are both female- (male-) dominated occupations with exactly equal gaps $|s^f_i - s^m_i| = |s^f_k - s^m_k|$, then a small shift of the female (male) labor force from occupation i to k should reduce (increase) the segregation index whenever $T_i/T < T_k/T$ $(T_i/T > T_k/T)$. o

A.8 represents a strong value judgement implying that, in a pair of female (male) occupations, it is more desirable to reduce the male-female ratio in the smaller one. The justification offered by Kakwani (1994) is that small occupations are generally among the higher paid ones. Therefore, gaps among them should be given larger weights.

Axiom 9: (Kakwani 1994) If i and k are both female- (male-) dominated occupations with size $T_i = T_k$, then a small shift of the female (male) labor force from occupation i to k should reduce (increase) the segregation index if $|s^f_i - s^m_i| > |s^f_k - s^m_k|$ $\left(|s^f_i - s^m_i| < |s^f_k - s^m_k|\right)$. o

Several contributions in the literature have emphasized the importance of basic aggregation properties. In this context, the simplest property that an index of segregation must satisfy is that a group with no members should have no effect on segregation. Consequently, one can delete occupations that contain no people without affecting measured segregation.

Axiom 10: (*Zero Member Independence*, Hutchens 2001). Let ($\mathbf{f}$, F, $\mathbf{m}$, M, $\mathbf{t}$, T) and ($\mathbf{f}'$, F, $\mathbf{m}'$, M, $\mathbf{t}'$, T) be identical except that ($\mathbf{f}'$, F, $\mathbf{m}'$, M, $\mathbf{t}'$, T) includes an occupation J + 1 with no members, $T_{J+1} = 0$, that is excluded from ($\mathbf{f}$, F, $\mathbf{m}$, M, $\mathbf{t}$, T). Then $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = \theta(\mathbf{f}', F, \mathbf{m}', M, \mathbf{t}', T)$. o

For the next property, it is useful to introduce the notion of a proportional division, an operation that divides an existing occupation into several new ones so that the gender ratio of female to male workers in the new occupations is equal to the original (predivision) ratio.

Definition 2: (Hutchens 2001) Let N be an integer. A distribution (**f'**, F, **m'**, M, **t'**, T) is said to be obtained from (**f**, F, **m**, M, **t**, T) through *a proportional division* of, say, occupation J, into N + 1 new ones, if $F'_j = F_j$ and $M'_j = M_j$ for all j ≠ J, and $F'_i = F_i/(N + 1)$ and $M'_i = M_i/(N + 1)$, so that $r'_i = r_i$ for all i = J, J + 1,…, J + N. o

The next axiom requires that an index be unaffected by the division of an occupation into units with identical segregation patterns. As pointed out by James and Tauber (1985), this principle has no analogue in the literature on income inequality measurement. It allows the comparison of economies with a different number of occupations, once the numbers are artificially equalized by a suitable division or combination of occupations.

Axiom 11: (*Organizational Equivalence*, James and Taeuber 1985, or *Insensitivity to Proportional Divisions*, Hutchens 2001) Let (**f'**, F, **m'**, M, **t'**, T) be obtained from a proportional division of an occupation of (**f**, F, **m**, M, **t**, T). Then θ(**f'**, F, **m'**, M, **t'**, T) = θ(**f**, F, **m**, M, **t**, T). o

Finally, in many contexts we are interested not only in the extent of gender segregation, but also in the actual pattern that characterizes this phenomenon in each occupation. Similarly, it may be useful to measure the contribution of each occupation, or a subset of them, to overall gender segregation. To formalize this idea, assume that the relevant information about gender segregation in each occupation j can be described by the 6-tuple (F_j, F, M_j, M, T_j, T) where, as before, $F = \Sigma_j F_j$, $M = \Sigma_j M_j$ and $T = \Sigma_j T_j$. A local index of gender segregation in that occupation, θ_j, will be a real valued function $\theta_j = \theta_j(F_j, F, M_j, M, T_j, T)$ satisfying A.1, A.2 and A.4. Now it is possible to state:

Axiom 12: (*Additivity*) For any ($\mathbf{f}$, F, $\mathbf{m}$, M, $\mathbf{t}$, T), $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = \Sigma_j \alpha_j \theta_j(F_j, F, M_j, M, T_j, T)$, where $\alpha_j \geq 0$ for all j, and $\Sigma_j \alpha_j = 1$. **o**

So far, the notion of segregation used refers to a situation in which the vectors $\boldsymbol{s}^{\mathrm{f}}$ and $\boldsymbol{s}^{\mathrm{m}}$ differ. However, segregation can also be said to exist when the female shares w_j differ across occupations, as in the entropy measure first proposed by Theil and Finizza (1971), or when it is the gender ratios r_j that differ across occupations, as in the index first suggested in Charles (1992). However, since $w_j \neq w_k$ for any $j, k \in \{1,\ldots, J\}$ if and only if $r_j \neq r_k$, these two notions need not be treated separately. In any case, it is easy to see that all axioms presented in terms of the vectors $\boldsymbol{s}^{\mathrm{f}}$ and $\boldsymbol{s}^{\mathrm{m}}$ (A.1, and A.6 to A.9), as well as Definition 1 can be equally written in terms of the vector(s) $\boldsymbol{w}$ (or $\boldsymbol{r}$).

II. 3. Invariance Axioms

In the literature on income inequality, it is customary to distinguish between indices that focus on income differences and indices that focus on income shares (see Kolm, 1999). In the first case, the measure of income inequality is invariant to equal additions to all incomes (translation invariance), and indices are referred to as absolute indices. In the second case, income inequality is not affected by proportional changes in all incomes (scale invariance), and indices are referred to as relative indices. Scale and translation invariance correspond to two particular inequality views so that the choice among them is normative and depends on value judgements.

In the segregation literature, most indices entail a relative view in which relative magnitudes are all that matters. Formally:

Axiom 13: (*Size Invariance*, James and Taeuber, 1985) Let ($\mathbf{f}'$, F', $\mathbf{m}'$, M', $\mathbf{t}'$, T') = ($l\mathbf{f}$, lF, $l\mathbf{m}$, lM, $l\mathbf{t}$, lT) where l is a positive scalar. Then $\theta(\mathbf{f}', F', \mathbf{m}', M', \mathbf{t}', T') = \theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T)$.
This axiom resembles the Population Principle in the income inequality literature according to which replications of the income distribution do not alter income inequality. Clearly, under A.13, all relative magnitudes - s^{f}, s^{m}, and s^{t}, $\mathbf{w}$, $(1 - \mathbf{w})$, $\mathbf{r}$, W, (1 – W), and R- remain constant.[9]

Beyond this fundamental distinction between relative and absolute views of segregation, in the empirical literature on gender segregation comparisons it has been noticed that both the overall gender composition of employment, W and $(1 - W)$, as well as the distribution of the employed population across occupations, $s^{\mathrm{t}} = (T_1/T, \ldots, T_J/T)$, typically change over time and/or space. Consider, for instance, intertemporal comparisons during periods characterized by increased female participation in the labor market and/or a drastic decline in the agricultural sector. Likewise, cross-country comparisons of gender segregation would be sensitive to differences in the overall shares of employment by gender and the distribution of employment by occupation. Consequently, it has been forcefully argued that rigorous comparisons of segregation in cross-country and time-series studies must be margin-

[9] For a study that focuses on translation invariant segregation indices that represent an absolute view of segregation, see Chakravarty and Silber (1992).

free, in the sense that they should be made independent of changes in W and $(1 - W)$ - composition invariance, James and Tauber (1985)- and changes in s^t -occupations invariance, Blackburn *et al.* (1993, 1995).[10] The following two axioms have been proposed to capture these ideas.

Axiom 14: (*Invariance 1*, *Homogeneity*, Hutchens 1991) Starting from ($\mathbf{f}$, F, $\mathbf{m}$, M, $\mathbf{t}$, T), let ($\mathbf{f}'$, F', $\mathbf{m}'$, M') = ($l\mathbf{f}$, lF, $d\mathbf{m}$, dM), so that $T' = lF + dM$ and $T_j = lF_j + dM_j$ for each j, where l and d are distinct, positive scalars. Then $\theta(\mathbf{f}', F', \mathbf{m}', M', \mathbf{t}', T') = \theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T)$. **o**

Under A.14, measured segregation remains constant in spite of the fact that W and $\mathbf{s}^t$ will vary. Thus, all segregation indices that satisfy A.14 are both composition *and* occupational invariance. Such indices will be referred to as *margin-free type 1* indices. The only relative magnitudes that must remain constant under the conditions of A.14 are the female and the male distributions across occupations, $\mathbf{s}^f$ and $\mathbf{s}^m$. Therefore, these are the only relevant magnitudes in the domain of margin-free type 1 indices.

An interesting consequence of A.14 should be mentioned here. Consider a distribution of people across gender and occupations represented by ($\mathbf{f}$, F, $\mathbf{m}$, M, $\mathbf{t}$, T). Arrange the occupations according to the ratios (s^f_j/s^m_j) in ascending order. A segregation curve, first suggested by Duncan and Duncan (1955), represents the cumulative fraction of females (on the ordinate) and the cumulative fraction of males (on the abcissa) when occupations are so ordered. A segregation curve S($\mathbf{f}$, F, $\mathbf{m}$, M, $\mathbf{t}$, T) is said to dominate another S($\mathbf{f}'$, F', $\mathbf{m}'$, M', $\mathbf{t}'$,

[10] On the desirability of margin-free indices, see also Charles (1992), Charles and Grusky (1995) and Grusky and Charles (1998). However, Flückiger and Silber (1999, pp. 84-85) show their reservations regarding the notions of

T'), denoted by $S(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) \succ S(\mathbf{f'}, F', \mathbf{m'}, M', \mathbf{t'}, T')$, if it lies at no point below and at some point above the other. In this case, the distribution ($\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T$) is ranked as less segregated than ($\mathbf{f'}, F', \mathbf{m'}, M', \mathbf{t'}, T'$). Thus, just as with Lorenz curves in the income inequality literature, nonintersecting segregation curves provide an (incomplete) ranking of distributions of people across occupations. A segregation index θ is said to be *consistent* with the ranking of distributions according to segregation curves if for any two distributions ($\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T$) and ($\mathbf{f'}, F', \mathbf{m'}, M', \mathbf{t'}, T'$), $S(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) \succ S(\mathbf{f'}, F', \mathbf{m'}, M', \mathbf{t'}, T') \Leftrightarrow S(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) < S(\mathbf{f'}, F', \mathbf{m'}, M', \mathbf{t'}, T')$. Hutchens (1991, 2001) established the following result.[11]

Proposition 1. A segregation index θ is consistent with the ranking of distributions according to segregation curves if and only if it satisfies axioms A.3 (*Symmetry in Groups*), A.6 (*Movement between Groups*), A.11 (*Insensitivity to Proportional Divisions*) and A.14 (*Invariance 1*). o

Finally, consider the next invariance axiom.

Axiom 15: (*Invariance 2*, Watts, 1998a) Starting from ($\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T$), let $\mathbf{f'}$ and $\mathbf{m'}$ be such that $F'_j = l_jF_j$ and $M'_j = l_jM_j$, so that $T'_j = l_jT_j$, $F' = \Sigma_j\ l_jF_j$, $M' = \Sigma_j\ l_jM_j$, and $T' = \Sigma_j\ l_jT_j$, where l_j are positive scalars for all j, and $l_i \neq l_k$ for at least two occupations i and k. Then $\theta(\mathbf{f'}, F', \mathbf{m'}, M', \mathbf{t'}, T') = \theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T)$. o

Under A.15, measured segregation remains constant in spite of the fact that W and $\mathbf{s}^t$ will vary. Thus, all segregation indices that satisfy A.15 are both occupational *and* composition

---

composition and occupational invariance.

[11] On the other hand, it can be shown that A.14 and a certain version of the additivity axiom A.12 imply an *additivity in groups* axiom (P.7) suggested in Hutchens (2001).

invariance. Such indices will be referred to as *margin-free type 2* indices.[12] The only relative magnitudes that must remain constant under A.15 are the gender ratios **r** or the shares **w** and (**1 – w**). Therefore, these are the only relevant magnitudes in the domain of margin-free type 2 indices.

Finally, notice that both axioms A.14 and A.15 can be satisfied simultaneously. For instance, the segregation index introduced in Charles (1992) is both margin-free type 1 and type 2.

II. 4. The Multidimensional Case

Gender segregation has traditionally been associated with occupational segregation. However, a number of studies have shown that this unidimensional approach is too restrictive: other job and worker characteristics, such as industry, private or public sector, ethnic group, level of education, and labor market status, exhibit both trends and patterns of segregation which add to our understanding of occupational segregation.[13]

For the following axiom, consider situations in which workers with a given characteristic, say a three-digit occupation, can be classified in terms of a second characteristic, a two-digit occupation, but not *vice-versa*. This case is referred to as "a pair of one-way classification variables".

Assume that there are I two-digit occupations, indexed by $i = 1,\dots, I$, and that each two-

---

[12] Axiom A.14 was referred to as *composition invariance* by James and Tauber (1985) and Watts (1998a), while A.15 was designated as *occupational invariance* in Watts (1998a). We have chosen to call them *Invariance 1* and *2* because both axioms are alternative ways to obtain margin-free indices.

[13] See, for instance, Jacobs (1989a), Jacobsen (1994), Deutsch *et al.* (1994), Watts (1997a), Blau *et al.* (1998), and Mora and Ruiz-Castillo (2003a, 2003b, 2003c).

digit occupation i can be sub-classified into a three-digit occupation $j \in G_i$. Let F_{ij}, M_{ij} and $T_{ij} = F_{ij} + M_{ij}$ be the number of females, males, and people, respectively, in three-digit occupation j within two-digit group i, and let $\mathbf{f} = (F_{11}, F_{12},...,F_{IJ})$, $\mathbf{m} = (M_{11}, M_{12},..., M_{IJ})$ and $\mathbf{t} = (T_{11}, T_{12},...,T_{IJ})$. Let $\mathbf{f}^i = (F_{i1}, F_{i2},..., F_{iJ})$, $\mathbf{m}^i = (M_{i1}, M_{i2},..., M_{iJ})$ and $\mathbf{t}^i = (T_{i1}, T_{i2},..., T_{iJ})$ be, respectively, the gender and people's frequencies across three-digit occupations within occupation i. Let $F_i = \Sigma_{j \in G_i} F_{ij}$, $M_i = \Sigma_{j \in G_i} M_{ij}$ and $T_i = \Sigma_{j \in G_i} T_{ij}$ be the number of females, males and people in group i, and denote by $\mathbf{f}_i = (F_1, F_2, ..., F_I)$, $\mathbf{m}_i = (M_1, M_2,..., M_I)$ and $\mathbf{t}_i = (T_1, T_2,..., T_I)$ the aggregated gender and people's frequencies across major occupations. Finally, let $F = \Sigma_i F_i$, $M = \Sigma_i M_i$ and $T = \Sigma_i T_i$ be the overall number of females, males and people, respectively.

Several measures of segregation are then available in this situation: (i) an overall measure of segregation, $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T)$; (ii) a *between-group* measure of segregation, $\theta(\mathbf{f}_i, F, \mathbf{m}_i, M, \mathbf{t}_i, T)$, computed as if there is only segregation at the two-digit level, i.e. computed as if, for each i, $F_{i1} = ... = F_{iJ} = F_i$, $M_{i1} = ... = M_{iJ} = M_i$, and $T_{i1} = ... = T_{iJ} = T_i$; and (iii) a *within-group* measure of segregation $\theta^i(\mathbf{f}^i, F_i, \mathbf{m}^i, M_i, \mathbf{t}^i, T_i)$ for each i, computed as if there is only segregation at the three-digit level, i.e. computed as if $F_1 = ... = F_I = F/I$, $M_1 = ... = M_I = M/I$, and $T_1 = ... = T_I = T/I$. In this context, a convenient property is that the overall measure of gender segregation can be expressed as the sum of two components: a *between-group* term, which captures the gender segregation in two-digit occupations; plus a weighted sum of *within-group* terms, where each of them captures the gender segregation induced by three-digit occupations within each two-digit occupation.[14]

Axiom 16: (*Additive Decomposability*) There exist $\nu_i \geq 0$ for all i with $\Sigma_i \nu_i = 1$, so that $\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = \theta(\mathbf{f}_i, F, \mathbf{m}_i, M, \mathbf{t}_i, T) + \Sigma_i \nu_i\, \theta^i(\mathbf{f}^i, F_i, \mathbf{m}^i, M_i, \mathbf{t}^i, T_i)$. **o**

Next, consider situations in which individuals can be classified in terms of a first characteristic, say educational attainment, and/or in terms of a second characteristic, say occupation. This case is referred to as "a pair of two-way classification variables".[15] Take now $\theta(\mathbf{f}_i, F, \mathbf{m}_i, M, \mathbf{t}_i, T)$ and $\theta^i(\mathbf{f}^i, F_i, \mathbf{m}^i, M_i, \mathbf{t}^i, T_i)$ as measuring segregation between and within education characteristics, and define $\theta(\mathbf{f}_j, F, \mathbf{m}_j, M, \mathbf{t}_j, T)$ and $\theta^j(\mathbf{f}^j, F_j, \mathbf{m}^j, M_j, \mathbf{t}^j, T_j)$, as measures of segregation between and within occupations.

<u>Remark</u>. Given a segregation index θ satisfying A.16, it is easy to show that it possesses the following commutative property. There exist υ_i and η_j with $\upsilon_i \geq 0$, $\eta_j \geq 0$ for each i anf j, and $\Sigma_i \upsilon_i = \Sigma_j \eta_j = 1$, so that

$$\theta(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = \theta(\mathbf{f}_j, F, \mathbf{m}_j, M, \mathbf{t}_j, T) + \Sigma_j \eta_j\, \theta^j(\mathbf{f}^j, F_j, \mathbf{m}^j, M_j, \mathbf{t}^j, T_j)$$

$$= \theta(\mathbf{f}_i, F, \mathbf{m}_i, M, \mathbf{t}_i, T) + \Sigma_i \upsilon_i\, \theta^i(\mathbf{f}^i, F_i, \mathbf{m}^i, M_i, \mathbf{t}^i, T_i).$$

III. AN ENTROPY BASED INDEX OF SEGREGATION

III. 1. Definition and Motivation

[14] Notice the analogy between this property and the additive decomposability property originally suggested in the income inequality literature by Bourguignon (1978) and Shorrocks (1980).

[15] This paper only examines the case in which segregation takes places along two dimensions. However, the extension of these properties to more than two dimensions is straightforward. For an empirical study in which the non-student population of working age is classified according to human capital characteristics, labor market status, and occupations, see Mora and Ruiz-Castillo (2003c).

In information theory, $I_j = w_j \log(w_j/W) + (1 - w_j) \log((1 - w_j)/(1 - W))$ is known as the expected information of the message that transforms the proportions $(W, (1 - W))$ to a second set of proportions $(w_j, (1 - w_j))$. The value of this expected information is zero whenever the two sets of proportions are identical, it takes larger and larger positive values when the two sets are more different, and it is symetrical in $(w_j, (1 - w_j))$. Therefore, I_j can be interpreted as an index of local segregation in occupation *j* within the approach reviewed in the previous section.

A weighted average of these J indices of local segregation will constitute an additive index of segregation. The selection of the weights is an important issue. One possible option is to give the same weight to each occupation, thus ensuring that the index is occupational invariant. However, we agree with England (1981) when she states: "The weighted index has more intuitive appeal. Suppose that occupations that segregate more (or less) grow faster over time, putting a greater (or lesser) number of persons into segregated work. I prefer an index that reveals this increase (or decrease) in segregation over one that adjusts the change out because it resulted from a change in the relative size of occupations that segregate to different extents." Thus, the I_E index of overall segregation is defined by

$$I_E = \Sigma_j s^{\mathbf{t}}_j I_j. \tag{1}$$

That is to say, I_E is the weighted average of the information expectations, with weights proportional to the number of people in the occupations.[16]

[16] See Mora and Ruiz-Castillo (2003a) for details on the seminal contribution to this approach by Theil and

Naturally, this choice of weighting scheme facilitates the satisfaction of certain axioms, namely, additivity (A.12) and additive decomposability (A.16), but it makes the satisfaction of both invariance axioms (A.14 and A.15) impossible. However, the violation of invariance axioms does not preclude the possibility of sensible pairwise comparisons. As will be seen below, the fulfillment of additivity opens the way to a decomposable approach that permits isolating a margin-free term independent of changes in the overall occupational distribution and the female share.

The index I_E has also been motivated from a statistical point of view. From equation (1), it is straightforward to show that I_E can also be expressed as:

$$I_E = W \Sigma_j \, s^f{}_j \log\left(s^f{}_j / s^t{}_j\right) + (1 - W) \Sigma_j \, s^m{}_j \log\left(s^m{}_j / s^t{}_j\right).$$

After some simple algebraic transformation, the exponential of I_E may be written as:

$$\exp I_E = \prod_j \left\{ s^f{}_j / (s^t{}_j \, W) \right\}^{(Fj/T)} \left\{ s^m{}_j / (s^t{}_j \, (1 - W)) \right\}^{(Mj/T)}.$$

As pointed out by Flückiger and Silber (1999, pp. 69), "the expression within the [first] curled brackets (...) is, in fact, equal to the *posterior probability* of having [F_j] individuals over what would be the *prior probability* of having such a number of individuals if one assumed independence between occupations and gender". The expression within the second curled brackets admits a similar interpretation. Finally, in the context of district versus school's racial composition, Zoloth (1974, pp 14-16 and 1976) shows that Theil and Finizza's index of segregation can also be interpreted as "a statistical measure of association, highly analogous to

Finizza (1971) and Fuchs (1975). For a different segregation index also related to the concept of entropy, see a squared coefficient of correlation, between the school that an individual student attends and the minority/non-minority status of that student".

III. 2. Basic Axioms

The index I_E satisfies *Complete Integration* (A.1) because if $s^f_j = s^m_j$ for all j, then $s^f_j = s^t_j$ and $s^m_j = s^t_j$, so that, from equation (2), $I_E = 0$. *Symmetry in Types* (A.3), *Symmetry in Groups* (A.4) and *Additivity* (A.12) follow directly from the definition of I_E.

I_E also fulfills *Complete Segregation* (A.2). Theil and Finizza (1971) show that I_E equals E - μ, where $E = W \log(1/W) + (1 - W) \log(1/(1 - W))$, $\mu = \Sigma_j s^t_j E_j$, and $E_j = w_j \log(1/ w_j) + (1 - w_j) \log(1/(1 - w_j))$.[17] Notice that E_j takes its minimum value, equal to 0, when $w_j = 0$. Otherwise, E_j is positive and reaches its maximum value, equal to log 2, when $w_j = 1/2$. To normalize E_j between 0 and 1, from here on it is assumed that all logarithms are in base 2. The same argument applies to E, which is also normalized between the unit interval. Now, if $w_j \in \{0,1\}$ for all j, then $E_j = 0$ for all j and $\mu = 0$, so that $I_E = E$. Given that μ is non-negative, I_E is bounded from above by E, which is itself bounded by 1. Therefore, I_E can only take values in the interval $[0, E] \subset [0, 1]$, and the index reaches its maximum when there is complete segregation.

To verify that I_E satisfies A.5 to A.9, it is useful to compute the marginal effect on I_E of

Hutchens (1991) and the discussion in Flückiger and Silber (1999).

[17] E and E_j are the entropy of a distribution with proportions (W, (1 – W)) and (w, (1 – w $_j$)), respectively. They measure the gender mix in the overall population and in occupation j, respectively.

an infinitesimal shift of the female population from occupation i to occupation k: $dF_k = - dF_i > 0$. From (1), we have that:

$$dI_E = \left\{ \partial[T_k\, I_k]/\partial F_k - \partial[\, T_i\, I_i]/\partial F_i \right\} dF_k/T. \tag{2}$$

For any occupation j:

$$\partial[T_j\, I_j]/\partial F_j = I_j + T_j \left(\partial\, I_j/\partial w_j\right)\left(\partial w_j/\partial F_j\right),$$

where $\partial I_j/\partial F_j = \log(w_j/W) - \log((1\text{-}w_j)/(1\text{-}W))$ and $\partial w_j/\partial F_j = (1 - w_j)/T_j$, so that:

$$\partial[T_j\, I_j]/\partial F_j = \log(w_j/W). \tag{3}$$

Applying equation (3) to equation (2), it is seen after some manipulation that:

$$dI_E = \log(w_k/w_i)\, dF_k/T. \tag{4}$$

For I_E, the *Principle of Transfers* (A.5) follows directly from equation (4) and the fact that in a female dominated occupation, say i, $w_i > W$, whilst in a male dominated occupation, say k, $w_k < W$, so that $w_i > w_k$ and $dI_E < 0$. Of course, the decrease in the segregation index will take place as long as $w_i > w_k$, so the transfer does not have to occur between a female and a male dominated occupation.

To show that I_E satisfies *Movement between Groups* (A.6), note that given Equation (4), if $w_k > w_i$, then $dI_E > 0$ for a sufficiently small change $dF_k = - dF_i$. However, the condition for $dI_E > 0$, i.e. $w_k > w_i$, will always be met after any disequalizing change and, therefore, $dI_E > 0$ for any feasible discrete change, i.e. for any $0 < d \leq F_i$. Thus, A.6 is satisfied by index I_E. Since $w'_k > w_k > w_i > w'_i$, it is straightforward to see by a similar argument that I_E satisfies *Increasing Returns to Movement Between Groups* (A.7).

To show that I_E fullfils A.8, it is enough to show that if occupations *i* and *k* have equal gaps and $s^t_i < s^t_k$ then $dI_E < 0$. First, note that if *i* and *k* have equal gaps, then T_i $(w_i - W) = T_k$ $(w_k - W)$. If $T_i < T_k$, then it follows that $w_i > w_k$. But then, from equation (4), $dI_E < 0$. Finally, the proof for A.9 follows straight from the fact that if $|s^f_i - s^m_i| > |s^f_k - s^m_k|$ and $T_i = T_k$, then $w_i > w_k$.

The proof that I_E satisfies *Zero Member Independence* (A.10) is immediate since $T_{J+1}/T = 0$. Finally, *Insensitivity to Proportional Divisions* (A.11) holds because I_E satisfies C*omplete Integration* (A.1) (see above) and *Additive Decomposability* (A.16) (see Mora and Ruiz-Castillo (2003a)).

III. 3. Invariance Axioms

It is easily seen that I_E satisfies *Size Invariance* (A.13), that is to say, I_E is a relative index of segregation. However, it is not margin-free because it depends both on the female share of the population and the distribution of people across occupations.

Since I_E is neither composition invariant nor occupational invariant, reasonable comparisons across countries or over time, as stressed by Watts (1998a), can only be made if there is a decomposition of the index that identifies a margin-free component which is not affected by changes in the gender composition of the population nor changes in the distribution across occupations.[18] For that purpose, it is useful to note that I_E can be written as:

$$I_E = \Sigma_j s^t_j \left\{ w_j \log (w_j) + (1 - w_j) \log (1 - w_j) \right\} + E. \tag{5}$$

Let $\tau = (\tau_1, ..., \tau_J)$, and φ be any J+1 real numbers such that (a) $0 \leq \tau_j \leq 1$ for any j, with $\Sigma_j \tau_j = 1$, and (b) $0 \leq \varphi \leq \log(2)$. Given the additive structure in equation (5), I_E can be decomposed in the following three components:

$$I_E = A1 + A2 + A3, \tag{6}$$

where:

$$A1 = \Sigma_j \tau_j \left\{ w_j \log(w_j) + (1 - w_j) \log(1 - w_j) \right\} + \varphi,$$

$$A2 = \Sigma_j \left\{ s^t_j - \tau_j \right\} \left\{ w_j \log(w_j) + (1 - w_j) \log(1 - w_j) \right\},$$

$$A3 = (E - \varphi).$$

Pairwise comparisons can be carried out using this decomposition. Consider as an illustration the comparison of segregation in two countries, denoted by superscripts *A* and *B*. By equation (6), the difference in segregation between country *A* and country *B*, $I_E^A - I_E^B$, can be expressed as:

$$I_E^A - I_E^B = GENCOM1 + OCUPMIX1 + GENMIX1 \tag{7}$$

where:

---

[18] This is the approach advocated *inter alia* by Blau and Hendricks (1979), Jonung (1984), Beller (1985), and Watts (1992, 1997a, 1998a).

$$\text{GENCOM1} = \Sigma_j\, \tau_j \Big\{ w_j^A \log(w_j^A) + (1 - w_j^A)\log(1 - w_j^A) - w_j^B \log(w_j^B) - (1 - w_j^B)\log(1 - w_j^B) \Big\}$$

$$\text{OCUPMIX1} = \Sigma_j \Big\{ s^{\mathbf{t}}{}_t^A - \tau_j \Big\} \Big\{ w_j^A \log(w_j^A) + (1 - w_j^A)\log(1 - w_j^A) \Big\} -$$

$$\Sigma_j \Big\{ s^{\mathbf{t}}{}_t^B - \tau_j \Big\} \Big\{ w_j^B \log(w_j^B) + (1 - w_j^B)\log(1 - w_j^B) \Big\}$$

$$\text{GENMIX1} = E^A - E^B.$$

Note that GENCOM1 is a margin-free term in the sense that it reports the difference in segregation if there were no differences by country in either the employment distribution across occupations or the entropy in the overall population. OCUPMIX1 gives the difference in segregation if there were no differences in either the gender composition by occupation or the entropy in the overall population. Finally, GENMIX1 shows the difference in segregation if there were no differences in either the employment distribution across occupations or the gender composition by occupation. The decomposition in equation (7) can be implemented once the values for $\{\tau , \varphi\}$ are chosen. Potential choices include, for example, $\{\tau, \varphi\} = \{s^{tA}, E^A\}$, $\{\tau , \varphi\} = \{s^{tB}, E^B\}$, or any linear combination of the two sets of values. For example, when $\{\tau , \varphi\} = \{s^{tB} , E^B\}$, then:

$$\text{GENCOM1} = \Sigma_j\, s^{\mathbf{t}}{}_j^B \Big\{ w_j^A \log(w_j^A / W^B) + (1 - w_j^A) \log((1 - w_j^A)/(1 - W^B)) \Big\} - I_{E^B} \quad (8)$$

$$\text{OCUPMIX1} = \Sigma_j \Big\{ s^{\mathbf{t}}{}_j^A - s^{\mathbf{t}}{}_j^B \Big\} \Big\{ w_j^A \log(w_j^A) + (1 - w_j^A) \log(1 - w_j^A) \Big\}$$

$$\text{GENMIX1} = E^A - E^B.$$

Therefore, when $\{\tau, \varphi\} = \{s^{tB}, E^B\}$, GENCOM1 has the mathematical structure of a difference between two segregation indices for countries *A* and *B* where the segregation index for *A* uses country *B*'s overall distribution across occupations and gender mix.

Alternatively I_E can be written as:

$$I_E = W \Sigma_j s^f_j \log(s^f_j) + (1 - W) \Sigma_j s^m_j \log(s^m_j) + IC, \qquad (9)$$

where $IC = \Sigma_j s^t_j \log(1/s^t_j)$ is Theil's index of concentration for the overall distribution of occupations. Let ω and γ be any 2 real numbers such that $0 \leq \omega \leq 1$ and $0 \leq \gamma \leq \log(J)$. Given equation (9), I_E can be decomposed in the following three components:

$$I_E = B1 + B2 + B3$$

where:

$$B1 = \omega \Sigma_j s^f_j \log(s^f_j) + (1 - \omega) \Sigma_j s^m_j \log(s^m_j) + \gamma,$$

$$B2 = (IC - \gamma),$$

$$B3 = (W - \omega) \Sigma_j \left\{s^f_j \log(s^f_j) + s^m_j \log(s^m_j)\right\}.$$

The difference in segregation between country *A* and country *B*, $I_E^A - I_E^B$, can now be expressed as:

$$I_E^A - I_E^B = GENCOM2 + OCUPMIX2 + GENMIX2 \qquad (10)$$

where:

$$GENCOM2 = \omega \Sigma_j s^{f,A}_j \log(s^{f,A}_j) + (1 - \omega) \Sigma_j s^{m,A}_j \log(s^{m,A}_j) -$$

$$\omega \Sigma_j s^{f,B}_j \log(s^{f,B}_j) - (1 - \omega) \Sigma_j s^{m,B}_j \log(s^{m,B}_j)$$

$$OCUPMIX2 = IC^A - IC^B,$$

$$\text{GENMIX2} = \Sigma_j \left\{W^A - \omega\right\}\left\{s^{f}{}_j{}^A \log(s^{f}{}_j{}^A) + s^{m}{}_j{}^A \log(s^{m}{}_j{}^A)\right\} -$$

$$\Sigma_j\left\{W^B - \omega\right\}\left\{s^{f}{}_j{}^B \log(s^{f}{}_j{}^B) + s^{m}{}_j{}^B \log(s^{m}{}_j{}^B)\right\}.$$

GENCOM2 is a margin-free term in the sense that it reports the change in segregation if there were no changes in the female and male shares or in Theil's index of concentration for the overall distribution of occupations. Different decompositions can be obtained by choosing ω and γ. Alternatives include, for example, $\{\omega, \gamma\} = \{W^A, IC^A\}$, $\{\omega, \gamma\} = \{W^B, IC^B\}$, or any linear combination of the two sets of values. For example, when $\{\omega, \gamma\} = \{W^B, IC^B\}$, then:

$$\text{GENCOM2} = W^B\, \Sigma_j\, s^{\mathbf{f}}{}_j{}^A \log(s^{\mathbf{f}}{}_j{}^A / s^{\mathbf{t}}{}_j{}^B) + (1 - W^B)\, \Sigma_j\, s^{\mathbf{m}}{}_j{}^A \log(s^{\mathbf{m}}{}_j{}^A / s^{\mathbf{t}}{}_j{}^B) - I_E{}^B \qquad (15)$$

$$\text{OCUPMIX2} = IC^A - IC^B,$$

$$\text{GENMIX2} = (W^A - W^B)\left\{s^{\mathbf{f}}{}_j{}^A \log(s^{\mathbf{f}}{}_j{}^A) + s^{\mathbf{m}}{}_j{}^A \log(s^{\mathbf{m}}{}_j{}^A)\right\}$$

In this case, GENCOM2 has the mathematical structure of a difference between two segregation indices for countries *A* and *B* where the segregation index for *A* uses country *B*'s overall female share and Theil's index of concentration for the overall distribution of occupations.

Karmel and MacLachlan (1988) point out that decompositions such as (9) and (11) will not, in general, isolate the changes in segregation due to the changes in the gender composition of individual occupations from changes in the overall distribution of employment across occupations and the female share. This is so because these factors are related to each other by the equality $W = \Sigma_j\, s^{t}{}_j\, w_j$; hence, they lack independence. A simple way to address this problem is by choosing $\{\tau, \varphi\}$ and $\{\omega, \gamma\}$ so that $\omega = \Sigma_j \tau_j w_j^A = \Sigma_j \tau_j w_j^B$. This can be done by first obtaining τ such that $\tau'(\boldsymbol{w}^A - \boldsymbol{w}^B) = 0$, $|\tau|=1$, and $\tau_j > 0$ for all j.[19] Then define $\omega = \Sigma_j \tau_j w_j^A$, $\varphi = \omega \log(1/\omega) + (1 - \omega) \log(1/(1 - \omega))$, and $\gamma = \Sigma_j \tau_j \log(1/\tau_j)$. Finally, choose either equation (7) or (10) to carry out the decomposition.[20]

III. 4. The Multidimensional Case

As already stated in subsection III.2, Mora and Ruiz-Castillo (2003a) show that I_E satisfies the additive decomposability property (A.16).[21] This property is useful to attack the following, rather classical problem. There is a potential bias due to small cell size (Blau *et al.*,1998): random allocations of individuals across occupations may generate high levels of gender segregation purely by chance. On the other hand, the use of more detailed categories leads to larger index values, since broader categories mask some of the segregation within them (England, 1981). Thus, it is interesting to study how far it is possible to aggregate an initial long list of occupations without reducing the gender segregation value too much. Herranz *et al.* (2003) propose an aggregation algorithm that uses I_E. The within-group term is

[19] Note that, since $\Sigma_j w_j^A = \Sigma_j w_j^B = 1$, and $w_j^A, w_j^B \in [0,1]$ for all j, there is always at least one vector satisfying all restrictions.

[20] Karmel and MacLachlan (1988) address this problem for their proposed index by implementing an iterative procedure to identify a joint distribution of occupations and gender for, say, country *A* that closely follows the same marginal distributions of country *B* and keeps the original association structure between gender and occupation (see also Watts, 1992). Their decomposition procedure, however, includes an interaction term.

[21] For his S_β family of indexes, Kakwani (1994) defines a between-group term, $\theta(\mathbf{f}_i, F, \mathbf{m}_i, M, \mathbf{t}_i, T)$, and a segregation index within a major occupation, $\theta^i(\mathbf{f}^i, F_i, \mathbf{m}^i, M_i, \mathbf{t}^i, T_i)$, but it does not establish the additive decomposability in the sense of A.16. For an alternative decomposition into three terms using the Gini-Segregation Index, see Silber (1989b), Boisso *et al.* (1994), Deutsch *et al.* (1994), and Sections 7.4 and 7.5 of Flückiger and Silber (1999). For the decomposition of the Karmel and MacLachlan segregation index into three terms see Borghans and Groot (1999).

identified as the error incurred in each step of the algorithm. Therefore, a reasonable stopping rule consists of selecting the furthest step for which the between group term is greater than or equal to the 1% bootstrapped lower bound for the original gender segregation value.

In the case of a pair of two way classification variables, the additive property of I_E, as well as its commutative property, has been repeatedly used in a number of recent applications (see Mora and Ruiz-Castillo, 2003a, 2003b, 2003c).

IV. STATISTICAL PROPERTIES

In this section, a statistical framework to test hypotheses on gender segregation using index I_E is developed. As Kakwani (1994) stresses, since segregation measures are estimated on the basis of sample observations, it is necessary to test whether the observed values are statistically significant. There are two approaches to this problem. First, it is possible to propose an index of segregation as a parameter or test in an econometric model. This is the direct approach taken by Charles (1992) and Kakwani (1994). Alternatively, bootstrap techniques can be used to compute standard errors in the segregation index, as in Deutsch *et al*. (1994) and Boisso *et al*. (1994). As bootstrap methods may fail if certain conditions are not met,[22] this approach requires verifying that these conditions are fulfilled by the index used in the empirical application.

IV. 1. A General Statistical Framework

Assume that a sample of T independent and identically distributed observations from

[22] See, for example, Davidson and Hinkley, (1999, pp. 38).

workers with information on their gender, $f_i \in \{0,1\}$ (one if worker i is female, zero if male), and occupation code, $z_i \in \{1,..., J\}$, is available. Consider the non-parametric model for the joint distribution:

$$\begin{aligned} &\Pr(f_i = 1, z_i = j) = \Omega_f, \ \Omega_{fj} \in [0,1], j = 1, ..., J, \\ &\Pr(f_i = 0, z_i = j) = \Omega_{mj}, \ \Omega_{mj} \in [0,1], j = 1, ..., J, \end{aligned} \tag{12}$$

where subindex f stands for female and subindex m stands for male and probabilities sum up to unity, $\Sigma_j(\Omega_{fj} + \Omega_{mj}) = 1$.

Consider the following two reparametrizations of this model based upon the multiplication rule. First, let $\Pr(z_i = j | f_i = 1) = \tau_{fj}$, $\Pr(z_i = j | f_i = 0) = \tau_{mj}$, and $\Pr(f_i = 1) = \omega$. The joint probability of gender and occupational category can be expressed as the product of the distribution of occupations conditional on gender, times the marginal distribution of gender:

$$\begin{aligned} &\Pr(f_i = 1, z_i = j) = \tau_{fj}\, \omega, \\ &\Pr(f_i = 0, z_i = j) = \tau_{mj}\, (1 - \omega) \end{aligned} \tag{13}$$

where $\omega \in [0,1]$ and $\tau_{fj}, \tau_{mj} \in [0,1]$, $\Sigma_j \tau_{fj} = \Sigma_j \tau_{mj} = 1$. Alternatively, let $\Pr(f_i = 1 | z_i = j) = \omega_j$ and $\Pr(z_i = j) = \tau_j$ for all $j = 1,...,J$. The joint distribution can also be expressed as the product of the probability of being female/male conditional on occupation times the marginal distribution of occupations:

$$\begin{aligned} &\Pr(f_i = 1, z_i = j) = \omega_j\, \tau_j \\ &\Pr(f_i = 0, z_i = j) = (1 - \omega_j)\, \tau_j \end{aligned} \tag{14}$$

where $\omega_j \in [0,1]$, and $\tau_j \in [0,1]$, $\Sigma_j \tau_{mj} = 1$. The crucial nonparametric feature in models (12), (13), and (14) is that parameters are free to vary along the categorical variables under consideration. This framework is more general than other models proposed in the literature. For example, Kakwani (1994) starts with a multinomial distribution for the observed frequencies $\mathbf{s}^f$. Then it assumes multivariate normality, which is a good approximation for large samples. In Charles (1992) and Charles and Grusky (1995), a pooled model with a log multiplicative specification for the expectation in the number of females and males in each cell is assumed.

IV. 2. Desirable Statistical Properties

Although models (12), (13), and (14) are observationally equivalent, they explore different, but not mutually exclusive, notions of statistical association between gender and occupation. The conditional probability in model (13) is a multinomial model for the occupation of either a female or a male worker. In contrast, the conditional probability in model (18) is a binomial model for the gender of a worker.

The most popular notion of segregation, that of gender differences in distributions across occupations, can be related to a testing procedure in model (13). Consider the hypothesis $\tau_{fj} = \tau_{mj} = \tau_j$, for all j, i.e. the probability that a worker is in occupation j is constant *regardless of the gender of the worker*. A test of this hypothesis can be seen as an intuitive statistical measure of segregation: the larger the value of the test, the less likely is the sample under absence of segregation.[23]

Denote by $L(\omega^{ML},\{\tau_{fj}^{ML}, \tau_{mj}^{ML}\}_j)$ the value of the log-likelihood of the sample $\{f_i, z_i\}_{i=1,\ldots,T}$ under model specification (13) at the Maximum Likelihood estimator, and let

$L(\omega^{ML},\{\tau_j^{ML}\}_j)$ be the log-likelihood for the restricted model where $\tau_{fj} = \tau_{mj} = \tau_j$, for all j. Finally, let ln() be the logarithm in base e. The index I_E satisfies the following:

Proposition 2: Let ξ_{LR1} be the log-likelihood ratio test for the equality of gender distributions across occupations in model (13),

$$\xi_{LR1} = -2 \, (L(\omega^{ML},\{\tau_j^{ML}\}_j) - L(\omega^{ML},\{ \tau_{fj}^{ML}, \tau_{mj}^{ML}\}_j).$$

Then: $\xi_{LR1} = 2 \, T \ln(2) \, I_E.$

o

The proof of Proposition 2 can be found in the Appendix. An alternative notion of segregation is provided by the analysis of the gender mix in each occupation as in model (14). Consider the hypothesis $\omega_j = \omega$, for all j, i.e. the probability that a randomly selected worker is female is constant, *regardless of the occupation of the worker*. A test of this hypothesis can also be seen as an intutitive statistical measure of segregation: the larger the test statistic, the less likely it is that absence of segregation is the true hypothesis. In the Appendix it is shown that the index I_E also satisfies the following:

Proposition 3: Let ξ_{LR2} be the log-likelihood ratio test for the equalities in the probabilities of being female across occupations in model (14),

$$\xi_{LR2} = -2 \, (L(\omega^{ML}, \{\tau_j^{ML}\}_j) - L(\{\omega_j^{ML},\tau_j^{ML}\}_j).$$

Then: $\xi_{LR2} = 2 \, T \ln(2) \, I_E.$

---

[23] This strategy for a statistical motivation of a segregation index has already been proposed by Kakwani (1994).

o

Therefore, the index I_E has two straightforward interpretations based on two log-likelihood tests. Under general conditions, this implies that a scaled version of I_E is asymptotically distributed as a $\chi^2_{(J-1)}$. It also follows that bootstrap methods to infer confidence intervals for small samples will be appropiate under general conditions not only for the index itself but also in *between/within* decompositions in the two-dimensional case and in decompositions in pairwise comparisons.

The results also highlight that, when individual data is available, a test on segregation in an unrestricted model naturally leads to an index which is not margin-free. The reason is that the test will exploit the statistical information embraced in the population distribution across occupations and the overall gender composition to characterize the null hypothesis scenario (i.e. absence of segregation), whilst in margin-free indeces all that matters is either the occupational distribution by gender or the gender mix by occupations.

V. CONCLUSIONS

This paper has presented an entropy based segregation index, I_E, that possesses a number of desirable properties. It satisfies all twelve basic axioms discussed during the last two decades for the single-dimensional case. It can be interpreted as two different log-likelihood tests so that bootstrap methods can be used to infer confidence intervals for small samples under general conditions. But it does not satisfy either of the two invariance axioms A.14 and A.15 that makes a segregation index compositional and occupational invariant. The failure to satisfy A.14 implies that the index is not consistent with the ordering provided by segregation curves. However, in pairwise comparisons it can be decomposed to isolate a margin-free term which reports the change in segregation holding the marginal distributions of gender and occupational category constant. Finally, in the two-dimensional case, it appears to be the only member of the class of relative indexes (satisfying axiom A.13) that possesses an *Additive Decomposability* property (A.16) analogous to the one that serves to characterize the family of generalized entropy indices in the income inequality literature.

How does I_E fare in relation to the remaining relative indexes of gender segregation either widely used or recently suggested? Consider first indices that are not embedded in a statistical framework and restrict the attention to the single-dimensional case.

1. The well known Dissimilarity Index is marginal-free of type 1 but, as pointed out in Zoloth (1976), James and Tauber (1985), and Hutchens (1991), it does not satisfy the strong versions of the Principle of Transfers, *Movement between Groups* (A.6) and *Increasing Returns to Movement between Groups* (A.7). A closely related index, originally suggested by Karmel and MacLachlan (1988), fails both invariance axioms A.14 and A.15 but it is decomposable into 4 terms, one of which is margin-free of type 1 and 2. However, it does not satisfy A.6 and A.7 either. This a serious drawback for a gender segregation index.

2. The Gini segregation index satisfies all basic axioms, except A.7 and *Additivity* (A.12), as well as the invariance axiom A.14. Therefore, it satisfies all the conditions of Proposition 1 (A.3, A.6, A.11, and A.14), so that it is consistent with the partial ordering obtained from segregation curves. It remains an interesting index, as has been extensively shown in Flückiger and Silber (1999).

3. Hutchens (2001) square root index

$$H(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = 1 - \Sigma_j \, (s^{\mathbf{f}}_j \, s^{\mathbf{m}}_j)^{1/2}$$

satisfies all the conditions of Proposition 1, as well as the remaining basic axioms. From this point of view, it deserves to begin to be used in empirical applications.

As indicated before, none of the above indices has been embedded in a statistical framework, a property that has been recently emphasized in the following two cases.

4. The logarithmic index suggested by Charles and Grusky (1995)

$$A(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = \exp \left[(1/J) \, [\Sigma_j \ln (r_j) - (1/J) \, \Sigma_j \ln (r_j)]^2\right]^{1/2}$$

is margin-free of type 1 and 2. However, the acceptance of a particular statistical model should depend on other desirable properties of the index obtained from it. In particular, as pointed out in Watts (1998a, 1998b), this index does violate *Organizational Equivalence* (A.11). Therefore, it is not consistent with the ordering provided by segregation curves. As indicated also by Watts (1998a, 1998b), this unweighted index is unduly influenced by extreme values caused by very low gender ratios that may characterize very small occupations. Moreover, if an occupation is completely segregated, with no (fe)male employees, the logarithm of the gender ratio $r_j = F_j/M_j$ is not defined.[24]

5. Like the I_E index advocated in this paper, Kakwani's (1994) preferred index

$$S_1(\mathbf{f}, F, \mathbf{m}, M, \mathbf{t}, T) = W \, (1 - W) \, \Sigma_j \, (s^{\mathbf{f}}_j - s^{\mathbf{m}}_j)^2 / s^{\mathbf{t}}_j$$

satisfies all basic axioms (except *Zero Member Independence*, A.10) but violates the invariance axioms A.14 and A.15. Although it has not yet been attempted, it would appear that there exists a decomposition of S_1 involving a margin-free term of the sort presented here in equations (7) and (10). This index deserves more applications beyond the only one known to Australia contained in Kakwani (1994).

It is important to emphasize that many of the above indexes can be extended to the two-dimensional case. However, none of them is additively decomposable in the sense of A.16 (for some alternatives, see note 21). It would appear that, together with some basic axioms, this property should serve to characterize the family of entropy based segregation indexes of which I_E is an interesting member – an exercise beyond the scope of this paper. In the meanwhile, it can safely be concluded that I_E is not inferior to any of its alternatives in the gender segregation literature.

---

[24] See, however, the reply by Grusky and Charles (1998).

APPENDIX

Proposition 2: Let ξ_{LR1} be the log-likelihood ratio test for the equality of gender distributions across occupations in model (17),

$$\xi_{LR1} = -2\ (\ L(\omega^{ML},\{\tau_j^{ML}\}_j) - L(\omega^{ML},\{\ \tau_{fj}^{ML},\ \tau_{mj}^{ML}\}_j).$$

Then: $\xi_{LR1} = 2\ T\ \ln(2)\ I_E.$

o

Proof: The only restrictions in model (17) are the boundness of ω, τ_{fj} and τ_{mj}, and also that $\Sigma_j\ \tau_{fj} = \Sigma_j\ \tau_{mj} = 1$. The likelihood function for worker *i* can be written as:

$$l_i\ (\omega,\{\tau_{fj},\ \tau_{mj}\}_j) = \prod_j \left\{(\tau_{fj})^{f_i\ Z_{ij}}\ (\tau_{mj})^{\ (1-f_i)Z_{ij}}\right\}\omega^{\ f_i}\ (1 - \omega)^{(1-\ f_i)}$$

where $Z_{ij} = Z(z_i = j)$ is the indicator function of worker *i* and occupation *j*. Therefore, the log-likelihood function of the sample is:

$$L(\omega,\{\tau_{fj},\ \tau_{mj}\}_j) = \Sigma_j\ T_j\ \left\{\ w_j \ln(\tau_{fj}) + (1 - w_j)\ \ln(\tau_{mj})\right\} + F \ln(\omega) + M \ln(1 - \omega).$$

Thus, $\omega^{ML} = W$, $\tau_{fj}^{ML} = F_j/F$, and $\tau_{mj}^{ML} = M_j/M$ for all *j*, and the value of the log likelihood function at the ML estimate is:

$$L(\omega^{ML},\{\tau_{fj}^{ML},\ \tau_{mj}^{ML}\}_j) = \Sigma_j\ T_j\ \left\{w_j \ln(F_j/F) + (1 - w_j)\ \ln(M_j/M)\right\}\ +\ F \ln(W) + M \ln(1-W)$$

In the restricted model, where $\tau_{fj} = \tau_{mj} = \tau_j$ for all *j*, the log-likelihood function for the entire sample is:

$$L(\omega,\{\tau_j\}_j) = \Sigma_j\left\{T_j \ln(\tau_j)\right\}\ + F \ln(\omega) + M \ln(1 - \omega).$$

The ML estimator for ω, ω^{ML}, is again the observed frequency of female workers in the sample: $\omega^{ML} = W$. In contrast to the unrestricted model, the ML estimator for τ_j now uses the marginal distribution of occupations for the entire population, $\tau_j^{ML} = T_j/T$ for all *j*. Therefore, the value of the log likelihood function of the restricted model at the ML estimate is:

$$L(\omega^{ML},\{\tau_j^{ML}\}_j) = \Sigma_j\ T_j\left\{w_j \ln(W) + (1 - w_j)\ \ln(1 - w_j) + \ln(T_j/T)\right\}.$$

Since

$$L(\omega^{ML},\{\tau_{fj}^{ML}, \tau_{mj}^{ML}\}_j) - L(\omega^{ML},\{\tau_j^{ML}\}_j) = \Sigma_j\, T_j \left\{w_j \ln(w_j/W) + (1 - w_j) \ln((1 - w_j)/(1 - W))\right\},$$

then the likelihood ratio test $\xi_{LR1} = 2\,T \ln(2)\, I_E$. ■

Proposition 3: Let ξ_{LR2} be the log-likelihood ratio test for the equalities in the probabilities of being female across occupations in model (18),

$$\xi_{LR2} = -2\,(L(\omega^{ML}, \{\tau_j^{ML}\}_j) - L(\{\omega_j^{ML},\tau_j^{ML}\}_j).$$

Then: $\xi_{LR2} = 2\,T \ln(2)\, I_E$.

o

Proof: There are no assumptions other than the boundness of ω_j and τ_j and that $\Sigma_j\, \tau_j = 1$. The log-likelihood function of the sample is:

$$L(\{\omega_j,\tau_j\}_j) = \Sigma_j\, T_j \left\{w_j \ln(\omega_j) + (1 - w_j) \ln(1 - \omega_j) + \ln(\tau_j)\right\}.$$

The ML estimator for ω_j, ω_j^{ML}, is simply the observed frequency of female workers within occupation j =1,...,J: $\omega_j^{ML} = w_j$. Also, $\tau_j^{ML} = T_j/T$. Therefore, the value of the log likelihood function at the ML estimator is:

$$L(\{\omega_j^{ML},\tau_j^{ML}\}_j) = \Sigma_j\, T_j\left\{w_j \ln(w_j) + (1 - w_j) \ln(1 - w_j) + \ln(T_j/T)\right\}.$$

In the restricted model, i.e. $\omega_j = \omega$ for all j, the logarithm of the likelihood function for worker i can be writen as:

$$L_i(\omega, \{\tau_j\}_j) = \Sigma_j\left\{f_i\, Z_{ij} \ln(\omega) + (1 - f_i)\, Z_{ij} \ln(1 - \omega) + Z_{ij}\ln(\tau_j)\right\}.$$

Thus, the log-likelihood function for the entire sample is:

$$L(\omega, \{\tau_j\}_j) = \Sigma_j T_j\left\{w_j \ln(\omega) + (1 - w_j) \ln(1 - \omega) + \ln(\tau_j)\right\} = F \ln(\omega) + M \ln(1 - \omega) + \Sigma_j T_j \ln(\tau_j).$$

The ML estimator for τ_j, τ_j^{ML}, is again the relative frequency of distribution j in the overall population, $\tau_j^{ML} = T_j/T$. In contrast to the unrestricted model, the estimator of ω, ω^{ML}, is now the observed frequency of female workers in the sample: $\omega^{ML} = W$. Therefore, the value of the log likelihood function at the ML estimate is:

$$L(\omega^{ML}, \{\tau_j^{ML}\}_j) = \Sigma_j\, T_j \left\{w_j \ln(W) + (1 - w_j) \ln(1 - W) + \ln(T_j/T)\right\}.$$

Since $L(\{\omega_j^{ML}, \tau_j^{ML}\}_j) - L(\omega^{ML}, \{\tau_j^{ML}\}_j) = \Sigma_j T_j \left\{w_j \ln(w_j/W) + (1 - w_j) \ln((1 - w_j)/(1 - W))\right\}$, the likelihood ratio test $\xi_{LR2} = 2\, T \ln(2)\, I_E$. ■

REFERENCES

Beller, A., 1985, Changes in the Sex Composition of U.S. Occupations, 1960-1981, Journal of Human Resources 20: 235-250.

Blackburn, R.M., Jarman, J., and J. Siltanen, 1993, The Analysis of Occupational Gender Segregation over Time and Place: Considerations of Measurement and Some New Evidence, Work, Employment and Society 7: 335-36.

Blackburn, R.M., Siltanen, J. and J. Jarman, 1995, The Measurement of Occupational Gender Segregation: Current Problems and a New Approach, Journal of the Royal Statistical Society A, Part 2 158: 319-331.

Blau, F., Simson, P. and D. Anderson, 1998, Continuing Progress? Trends in Occupational Segregation Over the 1970s and 1980s, Feminist Economics 4: 29-71.

Blau, F. and W. Hendricks, 1979, Occupational Segregation by Sex: Trends and Prospects, Journal of Human Resources 12: 197-210.

Boisso, D., Hayes, K., Hirschberg, J. and J. Silber, 1994, Occupational Segregation in the Multidimensional Case: Decomposition and Test of Significance, Journal of Econometrics 61: 161-171.

Borghans, L., and L. Groot, 1999, Educational Presorting and Occupational Segregation, Labour Economics 6: 375-395.

Bourguignon, F., 1979, Descomposable income inequality measures, Econometrica, 47: 901-920.

Butler, R.J., 1987, New Indices of Segregation, Economic Letters 24:359-363.

Chakravarty, S.R. and J. Silber, 1992, Employment Segregation Indices: An Axiomatic Characterization, in: W. Eichhorn, ed., Models and Measurement of Welfare and Inequality (Springer-Verlag, New York).

Charles, M., 1992, Cross-National Variation in Occupational Sex Segregation, American Sociological Review 57: 483-502.

Charles, M., 1998, Structure, Culture, and Sex Segregation in Europe, Research in Social Stratification and Mobility 16: 89-116.

Charles, M. and D. Grusky, 1995, Models for Describing the Underlying Structure of Sex Segregation, American Journal of Sociology 100: 931-971.

Cortese, C.F., Falk, R.F., and J.K. Cohen, 1976, Further Considerations on the

Methodological Analysis of Segregation Indices, American Sociological Review 41: 630-637.

Davison, A.C. and D.V. Hinkley, 1997, Bootstrap Methods and their Applications (Cambridge University Press, Cambridge).

Deutsch, J, Flückiger, Y. and J. Silber, 1994, Measuring Occupational Segregation, Journal of Econometrics 61: 133-146.

Duncan, O. and B. Duncan, 1955, A Methodological Analysis of Segregation Indexes, American Sociological Review 20: 210-217.

England, P., 1981, Assessing Trends in Occupational Sex Segregation, 1900-1976, in: I. Berg, ed., Sociological Perspectives on Labor Markets (Academic Press, New York).

Flückiger, Y. and J. Silber, 1999, The Measurement of Segregation in the Labor Force (Physica-Verlag, Heidelberg).

Fuchs, V., 1975, A Note on Sex Segregation in Professional Occupations, Explorations in Economic Research 2: 105-111.

Grusky, D.B. and M. Charles, 1998, The Past, Present, and Future of Sex Segregation Methodology, Demography 35: 497-504.

Herranz, N., Mora, R., and J. Ruiz-Castillo, 2003, An Algorithm to Reduce the Occupational Space in Gender Segregation Studies, forthcoming in Journal of Applied Econometrics.

Hutchens, R. M., 1991, Segregation Curves, Lorenz Curves and Inequality in the Distribution of People Across Occupations, Mathematical Social Sciences 21: 31-51.

Hutchens, R. M., 2001, Numerical Measures of Segregation: Desirable Properties and Their Implications, Mathematical Social Sciences 42: 13-29.

Jacobs, J., 1989a, Long-Term Trends in Occupational Segregation by Sex, American Journal of Sociology 95:160-173.

Jacobs, J., 1989b, Revolving Doors: Sex Segregation and Women's Careers (Stanford University Press, California).

Jacobsen, J., 1994, Trends in Work Force Sex Segregation, 1960-1990, Social Science Quarterly 75, 204-211.

James, D.R. and K.E. Taeuber, 1985, Measures of Segregation, in G. Schmid and R. Weitzel, eds., Sex Discrimination and Equal Opportunity: The Labor Market and Employment Policy, London, Gower Publishing Company.

Jonung, C., 1984, Patterns of Occupational Segregation by Sex in the Labor Market, in: N.B. Tuma, ed., Sociological Methodology (Jossey-Bass, San Francisco).

Kakwani, N.C., 1994, Segregation by Sex: Measurement and Hypothesis Testing, Research on Economic Inequality 5: 1-26.

Karmel, T. and M. MacLachlan, 1988, Occupational Sex Segregation: Increasing or Decreasing?, Economic Record 64:187-195.

Kolm, S.C., 1999, The Rational Foundations of Income Inequality Measures, in: J. Silber, ed., Handbook of Income Inequality Measurement (Kluwer Academic Publishers, , Dordretch).

Lewis, D.E., 1982, The Measurement of the Occupational and Industrial Segregation of Women, Journal of Industrial Relations 24: 406-423.

Moir, H., and J. Selby Smith, 1979, Industrial Segregation in the Australian Labour Market, Journal of Industrial Relations 21: 281-291.

Mora, R. and J. Ruiz-Castillo, 2003a, Additively Decomposable Segregation Indexes. The Case of Gender Segregation By Occupations and Human Capital Levels In Spain, forthcoming in Journal of Economic Inequality.

Mora, R. and J. Ruiz-Castillo, 2003b, Gender Segregation by Occupations in the Public and the Private Sectors. The Case of Spain In 1977 and 1992, Working Paper No. 03-16 Universidad Carlos III, Madrid.

Mora, R. and J. Ruiz-Castillo, 2003c, Gender Segregation: From Birth to Occupation, Working Paper No. 03-36, Economic Series 12, Universidad Carlos III, Madrid.

Peach, C., 1975, Urban Residential Segregation (Longman, London).

Shorrocks, A. F., 1980, The Class of Additively Decomposable Inequality Measures, Econometrica, 48: 613-625.

Shorrocks, A. F. and J. Foster, 1987, Transfer Sensitive Inequality Measures, Review of Economic Studies 54: 485-497.

Schwartz, J. and C. Winship, 1979, The Welfare Approach to Measuring Inequality, in: K.F. Schuessler, ed., Sociological Methodology (Jossey-Bass, San Francisco).

Silber, J., 1989a, On the Measurement of Employment Segregation, Economic Letters 30: 237-242.

Silber, J., 1989b, Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality, Review of Economics and Statistics LXXI: 107-115.

Silber, J., 1992, Occupational Segregation Indices in the Multidimensional Case: A Note, Economic Record 68: 276-277.

Siltanen, J., 1990, Social Change and the Measurement of Occupational Segregation by Sex: An Assessment of the Sex-Ratio Index, Work, Employment and Society, 4: 1-29.

Theil, H. and A.J. Finizza, 1971, A Note on the Measurement of Racial Integration of Schools by Means of Information Concepts, Journal of Mathematical Sociology 1: 187-194.

Watts, M., 1992, How Should Occupational Segregation Be Measured?, Work, Employment and Society 6: 475-487.

Watts, M., 1994, A Critique of Marginal Matching, Work, Employment and Society 8: 421-431.

Watts, M., 1997a, Multidimensional Indexes of Occupational Segregation, Evaluation Review 21: 461-482.

Watts, M., 1997b, The Measurement of Occupational Gender Segregation, Journal of the Royal Statistical Society A, 160: 141-145.

Watts, M., 1998a, Occupational Gender Segregation: Index Measurement and Econometric Modeling, Demography 35: 489-496.

Watts, M., 1998b, The Analysis of Sex Segregation: When is Index Measurement Not Index Measurement?, Demography 35: 505-508.

Zoloth, B.S., 1974, An Investigation of Alternative Measures of School Segregation, IRP Discussion Paper No. 229-74, University of Winsconsin-Madison.

Zoloth, B.S., 1976, Alternative Measures of School Segregation, Land Economics 52: 278-298.